

Abstract

Recently, the l_0 -least mean square (l_0 -LMS) algorithm has been proposed to identify sparse linear systems by employing a sparsity-promoting continuous function as an approximation of l_0 pseudonorm penalty. However, the performance of this algorithm is sensitive to the appropriate choice of the some parameter responsible for the zero-attracting intensity. The optimum choice for this parameter depends on the signal-to-noise ratio (SNR) prevailing in the system. Thus, it becomes difficult to fix a suitable value for this parameter, particularly in a situation where SNR fluctuates over time. In this work, we propose several adaptive combinations of differently parameterized l_0 -LMS to get an overall satisfactory performance independent of the SNR, and discuss some issues relevant to these combination structures. We also demonstrate an efficient partial update scheme which not only reduces the number of computations per iteration, but also achieves some interesting performance gain compared with the full update case. Then, we propose a new recursive least squares (RLS)-type rule to update the combining parameter more efficiently. Finally, we extend the combination of two filters to a combination of M number adaptive filters, which manifests further improvement for $M > 2$.

Adaptive Combination of l_0 -LMS Adaptive Filters for Sparse System Identification in Fluctuating Noise Power

Bijit K. Das, Mrityunjoy Chakraborty

May 11, 2016

1 Introduction

Exploiting sparsity of the identifiable system has been a celebrated topic in the last decade among adaptive filtering research community. However, it has a long and diverse history within and outside the adaptive signal processing community. Researchers in the field of network and acoustic echo cancellation have tried to take advantage of the sparse echo path models for improving the performance of the echo cancellation algorithms. Adaptive sparse system identification has also found applications in the field of sparse wireless multipath channel estimation and shallow underwater acoustic communication channel estimation. Starting from the adaptive delay filters and active tap detection based algorithms to the popular family of proportionate-type adaptive filters and recently proposed set theoretic adaptive filters, numerous attempts have been made to use the a priori knowledge about system sparsity.

Another family of sparsity-promoting norm regularized algorithms has gained immense popularity in recent years. Historically, the basis pursuit and other related methods have shown advantages of this approach. After the advent of sparse signal reconstruction techniques in the compressive sensing literatures, different sparsity-promoting norms have been borrowed by the adaptive filtering researchers.

Though the l_0 -pseudonorm which measures the sparsity by counting the number of nonzero elements in a vector can not be directly used for the regularization purpose since it is not a continuous function. Several approximations of this have been considered. The l_1 -norm or the absolute sum, the log-sum have been studied to a reasonable extent in this context, and have given birth to the algorithms like ZA-LMS and the RZA-LMS. Another approximation of l_0 -norm by some exponential function has also been proposed. This algorithm manifests excellent behaviour in terms

*B. K. Das and M. Chakraborty are with the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, INDIA (e.mail : bijitbijit@gmail.com; mrityun@ece.iitkgp.ernet.in).

of convergence speed and steady-state mean square deviation for proper choice of a parameter responsible for zero-attracting intensity.

In section *II*, we provide a brief review of the algorithm.

In section *III*, it has been shown that the choice of this parameter is sensitive to the signal-to-noise ratio (SNR) of the setup. An adaptive technique has been proposed to tackle this problem but with increasing complexity of the algorithm.

In section *IV*, we propose a convex combination of two differently parameterized l_0 -LMS adaptive filters to alleviate the sensitivity of selection of this parameter to some extent. We also demonstrate the simulation results to support the advantage of the proposed scheme.

In section *V*, we discuss a reduced-complexity partial update scheme for this combination. In this section, we also derive a new recursive least squares (RLS) type update rule for efficient adaptation of the combining parameter. The simulation results, we provide in this section, show how this update scheme proves to be a better alternative of the conventional gradient descent based update rule.

In section *VI*, we extend the techniques presented in the earlier sections to a more general combination of M number of adaptive filters.

2 Brief Review of the l_0 -LMS Algorithm

For deriving the l_0 -LMS algorithm, the cost function is modified as

$$L_0(n) = \frac{1}{2}e^2(n) + \gamma_2 \|\mathbf{w}(n)\|_0 \quad (1)$$

Considering that the l_0 -norm minimization is a Non-Polynomial (NP) hard problem, l_0 -norm is generally approximated by a continuous function. A popular approximation is

$$\|\mathbf{w}(n)\|_0 \approx \sum_{i=1}^L \left(1 - e^{-\beta|w_i(n)|}\right) \quad (2)$$

By minimizing the above cost function the gradient descent recursion for the i^{th} filter coefficient becomes

$$\begin{aligned} w_i(n+1) &= w_i(n) - \kappa\beta \text{sgn}\{w_i(n)\}e^{-\beta|w_i(n)|} \\ &\quad + \mu e(n)x(n-i+1) \\ (\forall 1 \leq i \leq L) \end{aligned} \quad (3)$$

To reduce the computational complexity, the first order Taylor series expansion of exponential functions is taken into consideration,

$$\begin{aligned} e^{-\beta|a|} &\approx 1 - \beta|a| \quad \text{if } |a| < \frac{1}{\beta} \\ &\approx 0 \quad \text{otherwise} \end{aligned} \quad (4)$$

It is to be noted that the above approximation is bounded to be positive because the exponential function is larger than zero.

Now, the approximated gradient descent recursion is

$$\begin{aligned} w_i(n+1) &= w_i(n) - \kappa \beta f_\beta(w_i(n)) + \mu e(n)x(n-i+1), \\ (\forall 1 \leq i \leq L), \end{aligned} \quad (5)$$

where

$$\begin{aligned} f_\beta(a) &= \beta^2 a + \beta, \quad \text{if } -\frac{1}{\beta} \leq a \leq 0; \\ &= \beta^2 a - \beta, \quad \text{if } 0 \leq a \leq \frac{1}{\beta}; \\ &= 0 \quad \text{elsewhere} \end{aligned}$$

3 The Dependence of the Optimum κ on the Signal-to-Noise Ratio (SNR)

In this section, we try to show how the optimum κ depends on the signal-to-noise ratio (SNR).

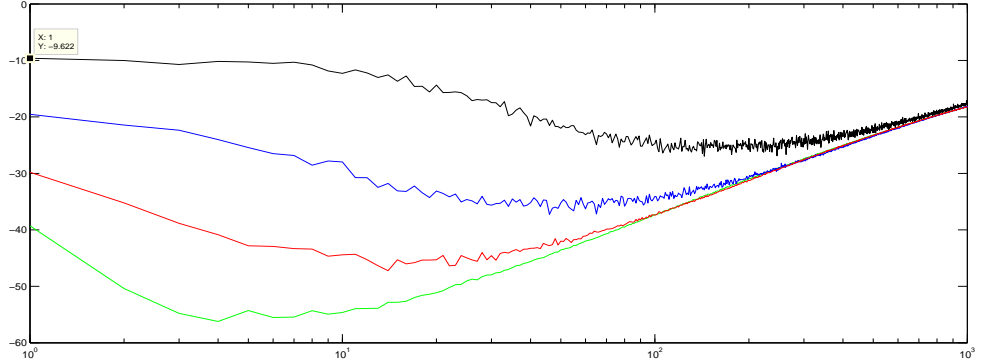


Figure 1: Steady-state MSD vs. κ for four SNR values [10dB (black), 20dB (blue), 30dB (red), 40dB (green)]

Experiment 1: We try to identify a sparse FIR system of length 128, which has only five nonzero coefficients (set as 1), and rest of them are zeros. We use a zero mean, unit variance white Gaussian random process as input to the unknown system as well as the l_0 -LMS adaptive filter. The parameters of the algorithm are set as: $\mu = 0.8$, . Matlab simulations are run for 15000 samples and steady-state MSD is calculated by averaging repeating the experiment over 100 iterations. The value of κ is varied from 0.000001 to 0.0001 with 1000 steps, and the steady-state MSD is

plotted against κ in Fig.1. Four different noise variances are chosen for the experiment. The black, blue, red and green curves correspond to 10dB, 20dB, 30dB, 40dB SNR-s respectively.

The optimum κ for which the minimum MSD is achieved is different for different SNRs. It explicitly implies that when SNR is likely to fluctuate over a wide range, a fixed single κ can not guarantee the best result for all different cases. This fact motivates us to propose a convex combination of two l_0 -LMS adaptive filters with two different κ values.

4 Proposed Adaptive Convex Combination of Two Differently Parameterized l_0 -LMS Adaptive Filters

In this section, we propose a convex combination of two l_0 -LMS adaptive filters. In this part of the paper, we have constrained ourselves to use the convex combination scheme in [1]. The outputs of the two l_0 -LMS adaptive filters are combined using the following rule:

$$y(n) = \lambda(n)y_1(n) + (1 - \lambda(n))y_2(n) \quad (6)$$

where $y_i(n)$ is the output of the i^{th} adaptive filter ($i = 1, 2$), and $y(n)$ is the combined output.

These two filters are identical in all aspects apart from their individual values of κ .

$$\lambda(n) = \frac{1}{1 + \exp(-a(n))} \quad (7)$$

The update rule for the parameter $a(n)$ is as follows:

$$a(n+1) = a(n) + \mu_c e(n)(y_1(n) - y_2(n))\lambda(n)(1 - \lambda(n)), \quad (8)$$

where $e(n) = d(n) - y(n)$.

It is also noteworthy that $a(n)$ is constrained within the range of ± 5 . This is a common practice to avoid $a(n)$ being stucked at any of the two extremes of its unconstrained range.

Experiment 2: The κ values are chosen as 5×10^{-5} and 5×10^{-6} . Three SNR values are chosen as 60dB, 40dB and 20dB. The simulation starts with 60dB SNR, at the 6000th iteration noise power increases to give a SNR of 40dB, and finally, it becomes 20dB at the 12000th iteration. The simulation stops at the 18000th iteration. The MSD-s are obtained by averaging 100 independent runs of this simulation. The μ_c is kept at 3000. All other parameters and variables are same as in the experiment 1. The instantaneous MSD-s are plotted against iteration index n in fig.2. The proposed combination (plotted in green) achieves the best steady-state MSDs for all SNR levels. The red and the black curves represent two l_0 -LMS filters (for $\kappa = 5 \times 10^{-5}$ and 5×10^{-6} respectively.)

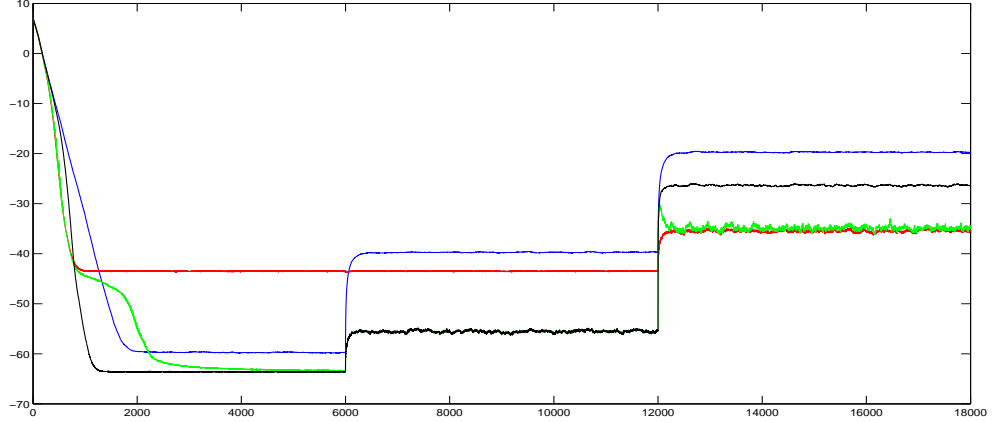


Figure 2: Instantaneous MSD Curves for the Proposed Combination (Green), l_0 -LMS (with $\kappa = 5 \times 10^{-5}$)(Red), l_0 -LMS (with $\kappa = 5 \times 10^{-6}$)(Black), LMS (Blue)

5 A Reduced Complexity Partial Update Scheme

One of the major drawbacks of the combination schemes, in general, is the increase in the number of computations per iteration by a multiple factor for deployment of multiple adaptive filters simultaneously. In this section, we demonstrate a partial update scheme for reducing the computational burden. We divide the set of all filter taps into two mutually exclusive subsets, named as *Odd* and *Even* for all odd and all even taps respectively. We update the filter taps of these individual subsets at alternative iterations. Now, one natural question that can arise is how we choose the subsets for individual component filters. Is it same for all the filters? Or, it is rather more beneficial to choose different subsets for different filters at a particular iteration. For the time being, we go for the second option without discussing the relative merits and demerits of this choice. We leave this discussion for the section VI.

5.1 A New RLS-type Update Rule for Adapting the Combining Parameter

We also derive a new update rule for $a(n)$ in the following. Let us first define the least-squares cost function $J_c(n)$ as

$$J_c(n) = \sum_{k=0}^n \beta^k e_c^2(n-k), \quad (10)$$

Table 1: The Partial Update Scheme with Mutually Exclusive Subsets for M Number of Filters

Initialization : $\mathbf{w}_k(0) = 0$ for each adaptive filter k . M number of diagonal matrices defined as \mathbf{S}_i ($i = 1, \dots, M$), for which $[\mathbf{S}_i]_{j,j} = 1$ for $j = i, i + M, i + 2M, \dots, i + \left\lfloor \frac{L}{M} \right\rfloor M$, and $= 0$ otherwise.

For every index n and every adaptive filter k , repeat

$$\begin{aligned} e_k(n) &= d_k(n) - \mathbf{w}_k^T(n) \mathbf{x}(n) \\ \mathbf{P}_k(n) &= \mathbf{S}_l \\ &\quad (l = \text{quotient}(k + n, M)) \\ \mathbf{w}_k(n+1) &= \mathbf{w}_k(n) + \mathbf{P}_k(n) [\mu_k \mathbf{x}_k(n) e_k(n) + \kappa_k \beta f_\beta(w_k(n))] \end{aligned} \quad (9)$$

where $e_c(n) = d(n) - y_c(n)$ is the error at the n^{th} instant for the combined filter.

$J_c(n)$ can be expanded by

$$J_c(n) = \sum_{k=0}^n \beta^k [d(n-k) - \lambda(n) \{y_1(n-k) - y_2(n-k)\} + y_2(n-k)]^2$$

The optimum value of $\lambda(n)$, $\lambda_{opt}(n)$ can be found as

$$\begin{aligned} \lambda_{opt}(n) &= \left[\sum_{k=0}^n \beta^k \{y_1(n-k) - y_2(n-k)\}^2 \right]^{-1} \left[\sum_{m=0}^n \beta^m \{y_1(n-m) - y_2(n-m)\} \{d(n-m) - y_2(n-m)\} \right] \\ &= [\mathbf{y}_d^T(n) \mathbf{B}(n) \mathbf{y}_d(n)]^{-1} [\mathbf{y}_d^T(n) \mathbf{B}(n) \mathbf{p}(n)], \end{aligned} \quad (12)$$

(13)

where $p(n) = d(n) - y_2(n)$, and $\mathbf{p}(n) = [p(n), p(n-1), \dots, p(0)]^T$.

Now, defining $r_{in}(n) = [\mathbf{y}_d^T(n) \mathbf{B}(n) \mathbf{y}_d(n)]^{-1}$, $y_d(n) = y_1(n) - y_2(n)$, $\mathbf{y}_d(n) = [y_d(n), y_d(n-1), \dots, y_d(0)]^T$, we get

$$\lambda_{opt}(n) = r_{in}(n) [\mathbf{y}_d^T(n) \mathbf{B}(n) \mathbf{p}(n)] \quad (14)$$

Now, to derive a recursive update for $\lambda_{opt}(n)$ using $y_d(n+1)$ and

$p(n+1)$, we write

$$\begin{aligned}
& \lambda_{opt}(n+1) \\
&= r_{in}(n+1)[\mathbf{y}_d^T(n+1)\mathbf{B}(n+1)\mathbf{p}(n+1)] \\
&= [\beta r_{in}^{-1}(n) + y_d^2(n+1)]^{-1} \{ \beta \mathbf{y}_d^T(n)\mathbf{B}(n)\mathbf{p}(n) + y_d(n+1)p(n+1) \} \\
&= \left\{ \frac{1}{\beta} r_{in}(n) - \frac{1}{\beta^2} y_d^2(n+1) r_{in}(n) [1 + y_d^2(n+1) r_{in}(n)]^{-1} r_{in}(n) \right\} \\
&\quad \left\{ \beta \mathbf{y}_d^T(n)\mathbf{B}(n)\mathbf{p}(n) + y_d(n+1)p(n+1) \right\} \\
&= \lambda_{opt}(n) - \frac{1}{\beta} y_d^2(n+1) \left[1 + \frac{1}{\beta} y_d^2(n+1) r_{in}(n) \right]^{-1} r_{in}(n) \lambda_{opt}(n) + \frac{1}{\beta} r_{in}(n) \\
&\quad \left[1 - \frac{1}{\beta} y_d^2(n+1) [1 + \frac{1}{\beta} y_d^2(n+1) r_{in}(n)]^{-1} r_{in}(n) \right] y_d(n+1) p(n+1) \\
&= \lambda_{opt}(n) - \frac{1}{\beta} y_d^2(n+1) \left[1 + \frac{1}{\beta} y_d^2(n+1) r_{in}(n) \right]^{-1} r_{in}(n) \lambda_{opt}(n) + \frac{1}{\beta} r_{in}(n) \\
&\quad \left[1 + \frac{1}{\beta} y_d^2(n+1) r_{in}(n) \right]^{-1} y_d(n+1) p(n+1) \\
&= \lambda_{opt}(n) + \frac{1}{\beta} r_{in}(n) \left[1 + \frac{1}{\beta} y_d^2(n+1) r_{in}(n) \right]^{-1} y_d(n+1) [p(n) - \lambda_{opt}(n) y_d(n)] \\
&= \lambda_{opt}(n) + \frac{1}{\beta} r_{in}(n) \left[1 + \frac{1}{\beta} y_d^2(n+1) r_{in}(n) \right]^{-1} y_d(n+1) [d(n) - \lambda_{opt}(n) y_1(n) - (1 - \lambda_{opt}(n)) y_2(n)] \\
&= \lambda_{opt}(n) + \frac{1}{\beta} r_{in}(n) \left[1 + \frac{1}{\beta} y_d^2(n+1) r_{in}(n) \right]^{-1} y_d(n+1) e_c(n) \\
&= \lambda_{opt}(n) + k(n) e_c(n), \tag{15}
\end{aligned}$$

where

$$k(n) = \frac{1}{\beta} r_{in}(n) \left[1 + \frac{1}{\beta} y_d^2(n+1) r_{in}(n) \right]^{-1} y_d(n+1) \tag{16}$$

But, since this update does not constrain $\lambda_{opt}(n)$ within 0 and 1, we update the corresponding variable $a_{opt}(n)$, and then compute $\lambda_{opt}(n)$ using the aforementioned rule. Since $\lambda_{opt}(n)$ is a monotonically increasing function $a_{opt}(n)$, the same incremental update has been used for $a_{opt}(n)$ as above.

$$a_{opt}(n+1) = a_{opt}(n) + k(n) e_c(n). \tag{17}$$

Experiment 3: Now, we repeat the experiment 2 with both the conventional gradient descent based update and the propose RLS-type update for the combiner separately, and we also incorporate the aforementioned partial update with mutually exclusive subsets scheme mentioned in the

table I for $M = 2$ case. We plot both the MSDs as well as the individual MSDs for the component filters and the simple LMS adaptive filter in fig.4 and fig.5 for SNR 20dB and 40dB respectively. For gradient descent based update, we plot two different curves for two different values of μ_c , i.e., 1000 and 10000 . For a particular SNR, one proves to be better than other. But, the RLS-type update shows its superiority by providing satisfactory behaviour in both situations without tuning its forgetting factor.

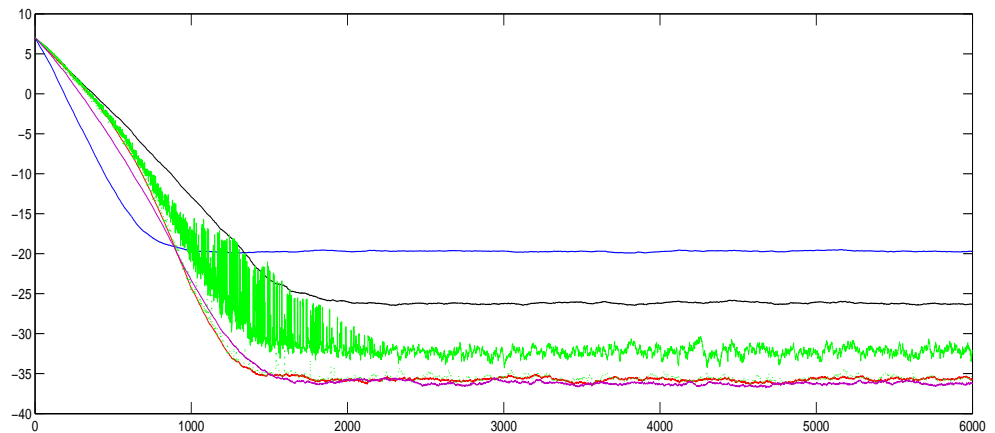


Figure 3: Instantaneous MSD Curves for the Combination (Green), l_0 -LMS (with $\kappa = 5 \times 10^{-5}$)(Red), l_0 -LMS (with $\kappa = 5 \times 10^{-6}$)(Black), LMS (Blue)

6 An Adaptive Convex Combination of M Number of Filters

In this section, we extend our work to a more general case of combining M adaptive filters so that a much wider range of SNR can be covered. We also modify the RLS-type combiner update rule for this general case. The derivation is skipped here. The rule is shown in the Table 2.

Experiment 4: We perform four separate experiments with three different SNR levels, i.e. 20dB, 40dB, 60dB and 40dB (with a different unknown system described below). In each case, the new combined adaptive filter ($M = 4$) is deployed to identify the aforementioned sparse unknown FIR system. The parameters of the filters remain same except κ chosen as 0, 1×10^{-6} , 1×10^{-5} and 5×10^{-5} . In each case, the robustness of the combination filter (in magenta) can be observed. In fig.8, a particular case has been studied where the unknown system is chosen as a near-sparse FIR system with all zero coefficients of the unknown system of the earlier experiments are replaced with very small non-zero values.

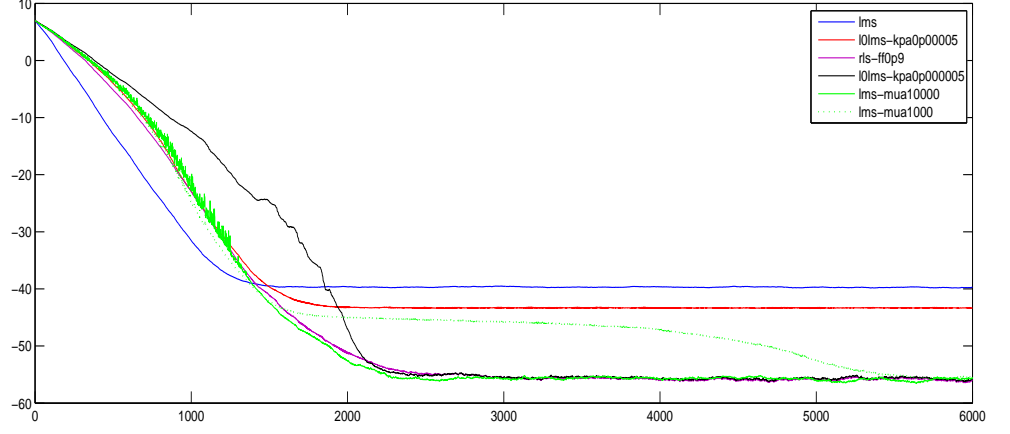


Figure 4: Instantaneous MSD Curves for the Proposed Combination (Green), l_0 -LMS (with $\kappa =$)(Red), l_0 -LMS (with $\kappa =$)(Black), LMS (Blue)

6.1 A Special Advantage of Using Partial Update with Mutually Exclusive Subsets

The next three figures demonstrate a special advantage of using the partial update scheme with mutually exclusive subsets compared with full update and PU with non-exclusive subsets. SNR level is 20dB and the unknown system is the near-sparse one in the last experiment. Here, though none of the component filters shows substantially better than the LMS, the combination manifests excellent performance for the PU with mutually exclusive subsets scheme [Fig.9]. Fig.10 and fig.11 are respectively for the full update and the PU with same subsets.

7 Partial Update with Uneven Mutually Exclusive Subsets

One major disadvantage of the partial update schemes mentioned in the last section is that it slows down the convergence of the individual adaptive filters as well as the combination by a fraction of M . In this section, we describe an uneven division of the filter tap indices into the subsets, and how it helps to alleviate the problem of slow convergence to some extent.

Let us here describe the uneven division of the subsets \mathbf{S}_l and how it differs from the one described in the table II. The rule is as follows:

$$\mathbf{S}_l[j, j] = 1 \text{ for } j = l : \Delta l : N \\ \text{and } = 0 \text{ elsewhere.}$$

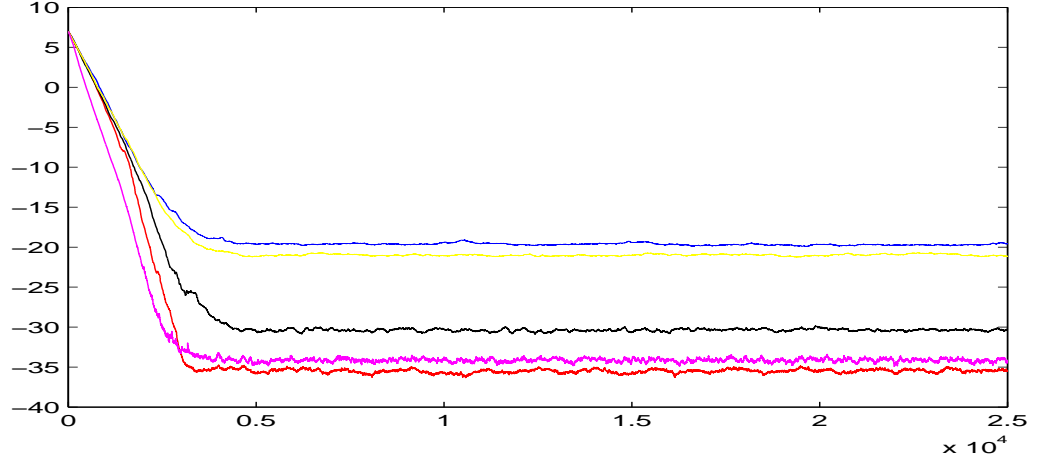


Figure 5: MSD vs. Iteration Index for Combination of $M = 4$ Adaptive Filters for $\text{SNR} = 20\text{dB}$

(for $l = 1$ to M . Δl is the smallest integer power of 2 greater or equal to l)

The presence of subsets with larger cardinalities help corresponding individual adaptive filters converge faster than those with smaller subsets. Thus, it leads to a better convergence behaviour for the overall combination as it is seen in the following figures [Fig. 10 -13]. The magenta curve (the combination) always follows the blue one (one with largest subset) first, then reconverges to the one with lesser steady-state m.s.d.

References

- [1] J. Radecki, Z. Zilic, and K. Radecka, "Echo Cancellation in IP Networks", *Proc. Forty-Fifth Midwest Symposium on Circuits and Systems*, vol. 2, 2002, pp. 219-222.
- [2] E. Hansler, "The Hands-free Telephone Problem - an Annotated Bibliography", *Signal Processing*, vol. 27, no. 3, pp. 259-271, Jun. 1992.
- [3] W. Schreiber, "Advanced Television Systems for Terrestrial Broadcasting", *Proc. IEEE*, vol. 83, no. 6, pp. 958-981, 1995.
- [4] W. Bajwa, J. Haupt, G. Raz and R. Nowak, "Compressed Channel Sensing", *Prof. IEEE CISS*, 2008, pp. 5-10.
- [5] M. Kocic, D. Brady and M. Stojanovic, "Sparse Equalization for Real-Time Digital Underwater Acoustic Communications", *Proc. IEEE OCEANS*, 1995, pp. 1417-1422.

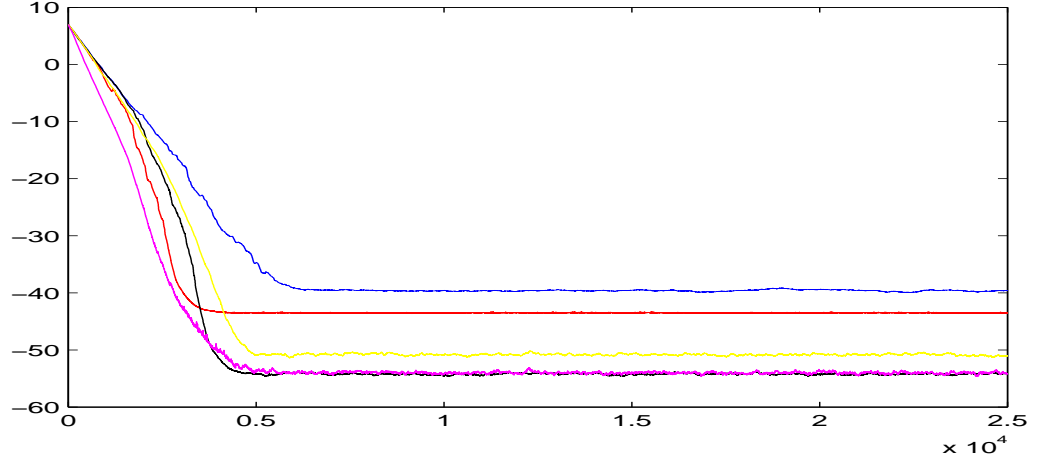


Figure 6: MSD vs. Iteration Index for Combination of $M = 4$ Adaptive Filters for SNR = 40dB

- [6] D. L. Duttweiler, "Proportionate Normalized Least Mean Square Adaptation in Echo Cancellers", *IEEE Trans. Speech Audio Processing*, vol. 8, no. 5, pp. 508-518, Sep., 2000
- [7] J. Benesty, S. L. Gay, "An Improved PNLMS Algorithm", *Proc. IEEE ICASSP*, vol. 2, pp. 1881-1884, May 2002, Orlando, Florida, USA.
- [8] H. Deng, and M. Doroslovacki, "Improving Convergence of the PNLMS Algorithm for Sparse Impulse Response Identification", *IEEE Signal Processing Letters*, vol. 12, No. 3, pp. 181-184, March 2005.
- [9] R. Baraniuk, "Compressive Sensing", *IEEE Signal Processing Magazine*, vol. 25, pp. 21-30, March 2007.
- [10] Y. Gu, Y. Chen and A. O. Hero, "Sparse LMS for System Identification", *Proc. IEEE ICASSP-2009*, April 2009, Taipei, Taiwan.
- [11] Y. Gu, J. Jin and S. Mei, " l_0 Norm Constraint LMS Algorithm for Sparse System Identification", *IEEE Signal Processing Letters*, vol. 16, No. 9, pp. 774-777, Sept. 2009.
- [12] B. K. Das and M. Chakraborty "Sparse adaptive filtering by an adaptive convex combination of the LMS and the ZA-LMS algorithms," *IEEE Trans. Circuits Syst.I, Reg.Papers*, Vol. 61, No. 5, May, 2014.
- [13] Dirk T. M. Slock, "On the Convergence Behavior of the LMS and the Normalized LMS algorithms", *IEEE Trans. on Signal Process.*, vol. 41, no. 9, pp. 2811-2825, September 1993.
- [14] J. Arenas-García, A. R. Figueiras-Vidal, and A. H. Sayed, "Mean-square performance of a convex combination of two adaptive filters", *IEEE Trans. Signal Proc.*, vol. 54, no. 3, pp. 1078-1090, March 2006.

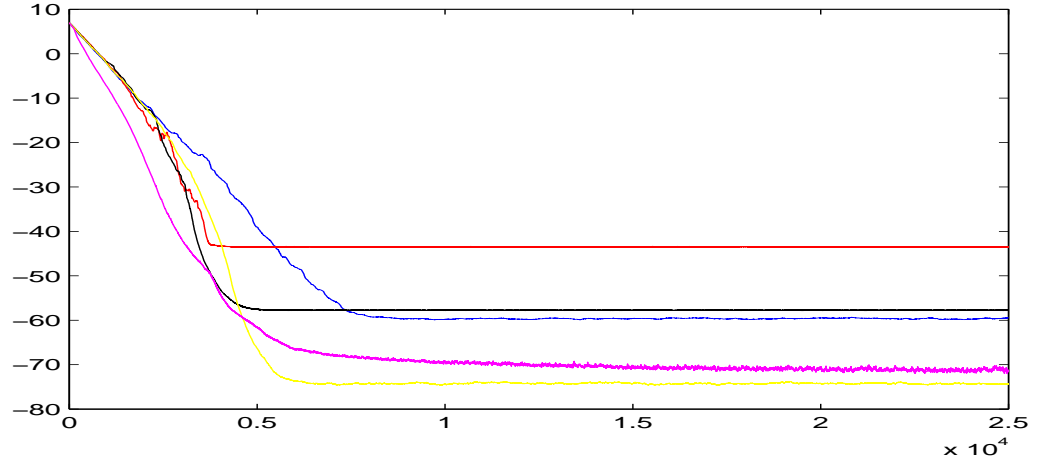


Figure 7: MSD vs. Iteration Index for Combination of $M = 4$ Adaptive Filters for $\text{SNR} = 60\text{dB}$

- [15] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes* (4th edition), New York: McGraw Hill, 2002.
- [16] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ: Prentice-Hall, 1986.

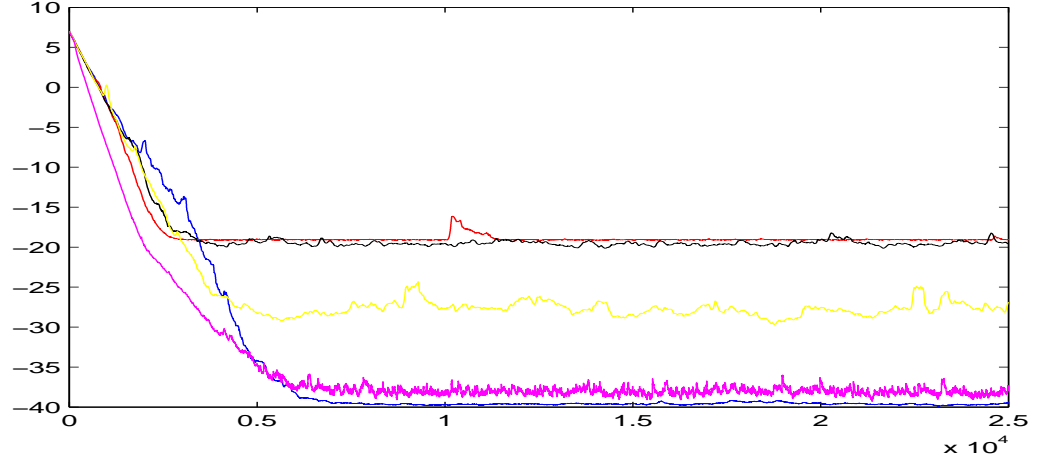


Figure 8: MSD vs. Iteration Index for Combination of $M = 4$ Adaptive Filters for $\text{SNR} = 40\text{dB}$ with near-sparse unknown system

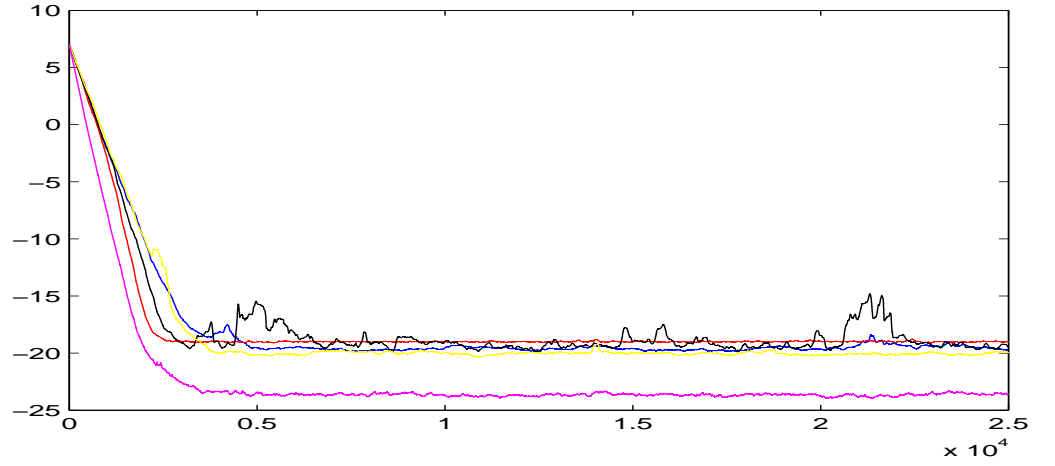


Figure 9: MSD vs. Iteration Index for Combination (PU with mutually exclusive subsets) of $M = 4$ Adaptive Filters for $\text{SNR} = 20\text{dB}$ with near-sparse unknown system

Table 2: The Partial Update Scheme with Mutually Exclusive Subsets for M Number of Filters

Initialization : $\mathbf{w}_k(0) = 0$ for each k .

$\mathbf{S}_l[j, j] = 1$ for $j = l : \left\lfloor \frac{N}{M} \right\rfloor : N$

and $= 0$ elsewhere.

(for $l = 1$ to M)

For every index n and every k^{th} adaptive filter, repeat

$$\begin{aligned}
 y_k(n) &= \mathbf{w}_k^T(n) \mathbf{u}(n) \\
 e_k(n) &= d_k(n) - y_k(n) \\
 \mathbf{w}_k(n+1) &= \mathbf{w}_k(n) + \mathbf{S}_{mod(n+k, M)} [\mu_k \mathbf{u}(n) e_k(n) + \kappa_k \beta f_\beta(\mathbf{w}_k(n))] \\
 \psi_k(n) &= \frac{\exp(\phi_k(n))}{\sum_{t=1}^M \exp(\phi_t(n))} \\
 y_c(n) &= \sum_{k=1}^M \psi_k(n) y_k(n) \\
 k_k(n) &= \frac{\lambda_f^{-1} p_k(n) (y_c(n) - y_k(n))}{1 + \lambda_f^{-1} p_k(n) (y_c(n) - y_k(n))^2} \\
 \phi_k(n+1) &= \phi_k(n) + k_k(n) e_k(n) \\
 p_k(n+1) &= \frac{1}{\lambda_f p_k^{-1}(n) + (y_c(n) - y_k(n))^2}
 \end{aligned} \tag{18}$$

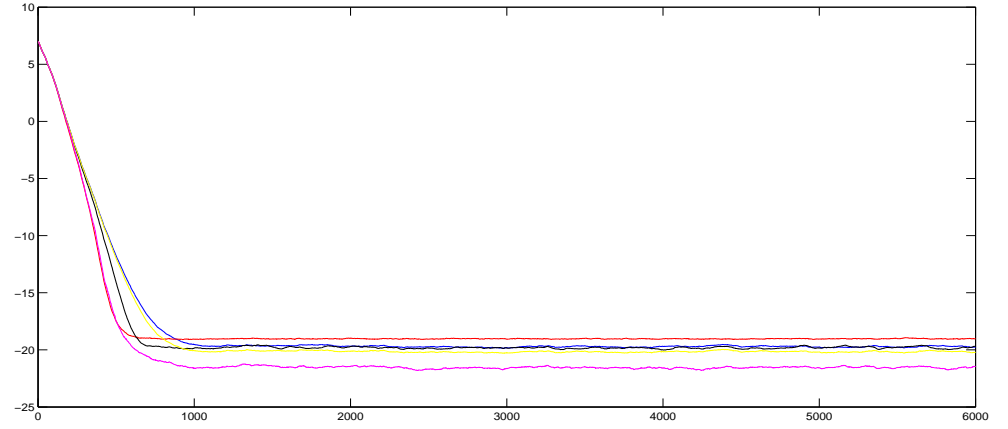


Figure 10: MSD vs. Iteration Index for Combination (full update) of $M = 4$ Adaptive Filters for SNR = 20dB with near-sparse unknown system

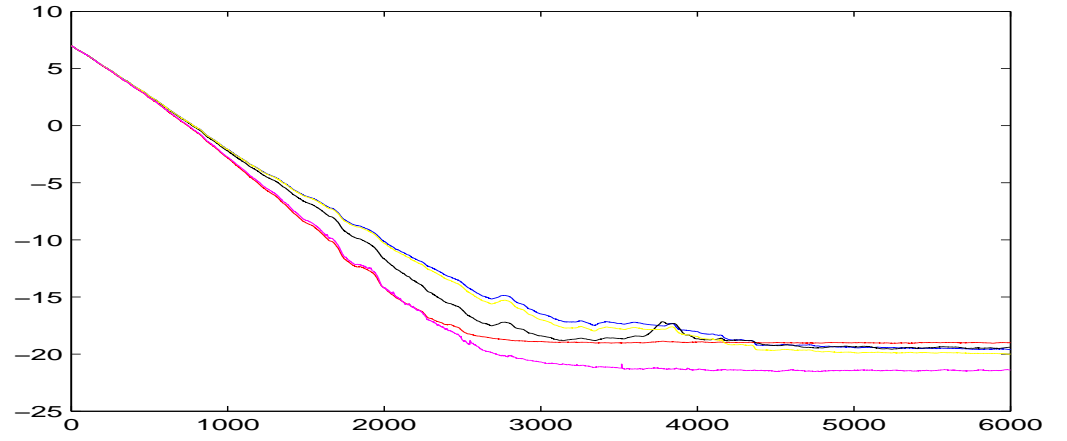


Figure 11: MSD vs. Iteration Index for Combination (PU with same subsets) of $M = 4$ Adaptive Filters for SNR = 20dB with near-sparse unknown system

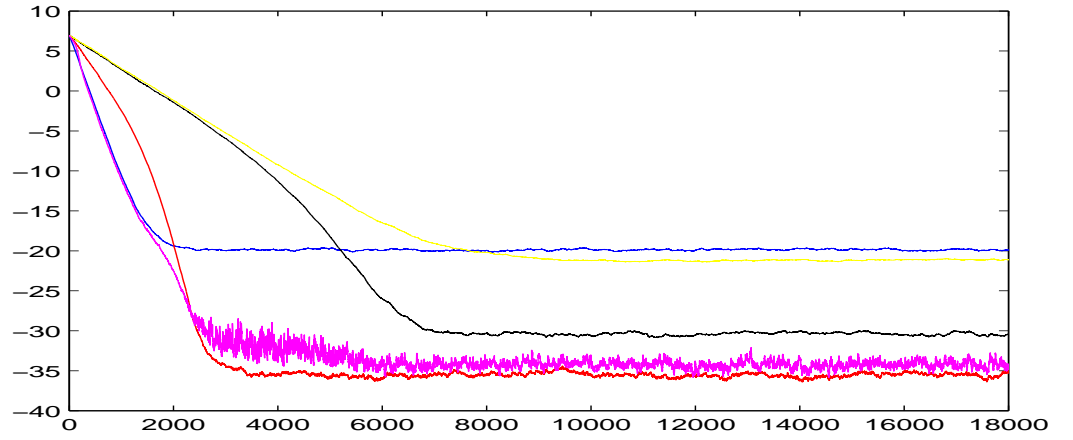


Figure 12: MSD vs. Iteration Index for Combination of $M = 4$ Adaptive Filters for $\text{SNR} = 20\text{dB}$

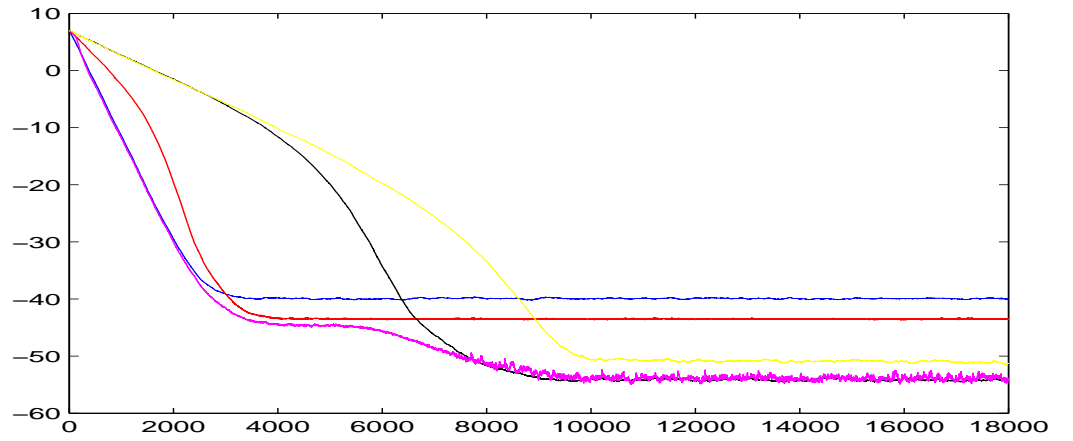


Figure 13: MSD vs. Iteration Index for Combination of $M = 4$ Adaptive Filters for $\text{SNR} = 40\text{dB}$

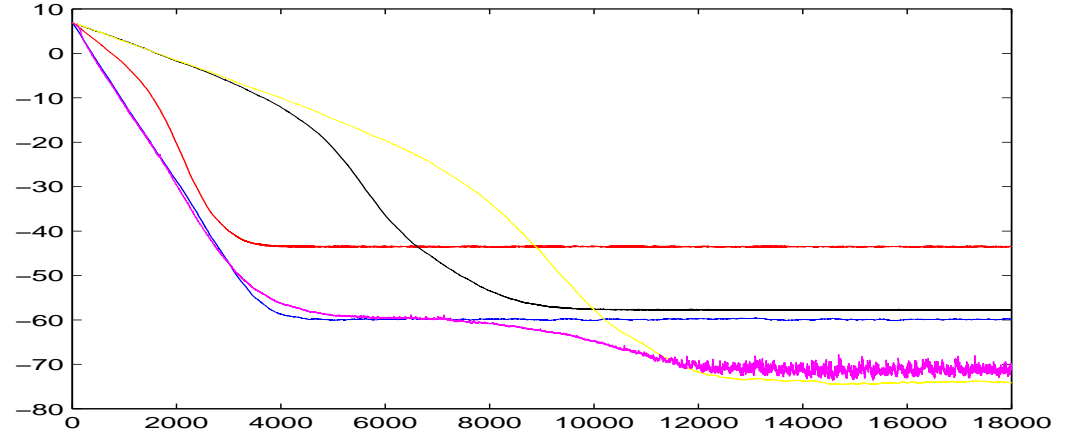


Figure 14: MSD vs. Iteration Index for Combination of $M = 4$ Adaptive Filters for $\text{SNR} = 60\text{dB}$

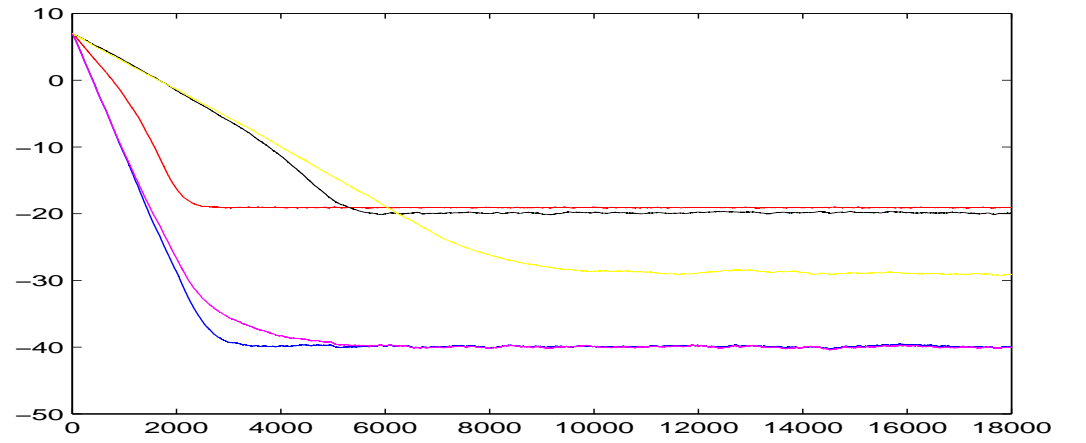


Figure 15: MSD vs. Iteration Index for Combination of $M = 4$ Adaptive Filters for $\text{SNR} = 40\text{dB}$ and the special \mathbf{w}_{opt}